



An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile

José Manuel Casado-Díaz, Lucas Martínez-Bernabéu & Francisco Rowe

To cite this article: José Manuel Casado-Díaz, Lucas Martínez-Bernabéu & Francisco Rowe (2017): An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile, *Spatial Economic Analysis*

To link to this article: <http://dx.doi.org/10.1080/17421772.2017.1273541>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 13 Feb 2017.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)



An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile

José Manuel Casado-Díaz^a, Lucas Martínez-Bernabéu^b and Francisco Rowe^{ib c}

ABSTRACT

An evolutionary approach to the delimitation of labour market areas: an empirical application for Chile. *Spatial Economic Analysis*. Labour market areas (LMAs) are argued to represent a more appropriate policy framework than administrative units for the analysis of spatial labour market activity. This article develops LMAs for Chile by applying an evolutionary computation approach. This innovative approach defines LMAs through an optimization process by maximization of internal cohesion, subject to restrictions of minimum levels of self-containment and population. To evaluate the appropriateness of the LMAs, comparative analyses are performed between alternative delimitations based on different parameter configurations of the proposed method versus administrative boundaries and the most widely used method for official LMA delimitation, the travel-to-work areas method.

KEYWORDS

evolutionary computation; labour market areas; travel-to-work areas; Chile

摘要

界定劳动市场范围的演化方法：智利的经验应用。*Spatial Economic Analysis*。在空间劳动市场活动的分析上，劳动市场范围（LMAs）被认为较行政单位更能呈现合适的政策架构。本文透过应用演化计算方法，为智利建立LMAs。此一创新方法，透过最大化内部凝聚力的最优化过程来定义LMAs，并受限于最低程度的自我封闭与人口。为了评估LMAs的适切性，本文根据提出方法相对于行政疆界的不同参数组合与“通勤范围方法”此一官方LMA界定最广为运用的方法，在不同的另类界定之间进行比较分析。

关键词

演化计算；劳动市场范围；通勤范围；智利


RÉSUMÉ

Approche évolutionnaire pour la délimitation de zones du marché du travail: une application empirique pour le Chili. *Spatial Economic Analysis*. Certains soutiennent que les zones du marché du travail constituent un

CONTACT

^a  jmcasado@ua.es

Department of Applied Economic Analysis and Institute of International Economics, University of Alicante, Alicante, Spain.

^b  lucas.martinez@ua.es

Institute of International Economics, University of Alicante, Alicante, Spain.

^c **(Corresponding author)**  F.Rowe-Gonzalez@liverpool.ac.uk

Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, Liverpool, UK.

cadre de politique plus approprié que les unités administratives pour l'analyse de l'activité spatiale du marché du travail. Cet article développe des zones du marché du travail pour le Chili en appliquant des principes de calcul évolutionnaires. Cette approche innovante définit des zones du marché du travail par le biais d'un processus d'optimisation par la maximisation de la cohésion interne, sous réserve de restrictions de niveaux maximum d'autonomie et de population. Afin d'évaluer la pertinence des zones du marché du travail, on effectue des analyses comparées entre délimitations alternatives d'après différentes configurations de paramètres entre la méthode proposée et les limites administratives, et la méthode la plus répandue pour la délimitation officielle des zones du marché du travail, la méthode des zones de déplacements à destination du lieu de travail.

MOTS-CLÉS

calcul évolutionnaire; zones du marché du travail; zones de déplacements à destination du lieu de travail; Chili

RESUMEN

Un enfoque evolutivo para la delimitación de las áreas del mercado laboral: aplicación empírica para Chile. *Spatial Economic Analysis*. Se argumenta que las áreas del mercado laboral (AML) representan un marco político más apropiado que las unidades administrativas para el análisis de la actividad del mercado laboral espacial. En este artículo se desarrollan las AML para Chile utilizando un enfoque de computación evolutiva. Este enfoque innovador define las AML mediante un proceso de optimización al maximizar la cohesión interna, sujeta a las restricciones de los niveles mínimos de autocontención y población. A fin de evaluar la idoneidad de las AML, se llevan a cabo análisis comparativos entre las delimitaciones alternativas basadas en diferentes configuraciones de parámetros del método propuesto frente a los límites administrativos y el método más utilizado para la delimitación oficial de las AML, el método de las cuencas de empleo.

PALABRAS CLAVES

computación evolutiva; áreas del mercado laboral; cuencas de empleo; Chile

JEL J61; O54; R12; R58

HISTORY Received 5 August 2014; accepted 9 December 2016

INTRODUCTION

Labour market areas (LMAs) constitute a type of *functional region* (Brown & Holmes, 1971) that captures the extent of commuting fields of residents and catchment areas of firms from a particular geographical area. They are, therefore, argued to provide a more appropriate spatial framework to capture the interplay between labour demand and supply than administrative geographical units (Goodman, 1970; Organisation for Economic Co-operation and Development (OECD), 2002; Smart, 1974). At the same time, spatial analyses based on administrative boundaries are often questioned due to the modifiable areal unit problem (MAUP) and measurement issues. These issues lead to spurious causal relationships, misguided policy recommendations and complex model specifications as a result of the presence of spatial spillovers (Newell, 2001). Although the use of modern spatial econometric and statistics techniques can alleviate these effects, having a more appropriate spatial framework is critical to capture accurately the geographical dynamics of labour market activity.

Over the last two decades, there has been a growing literature concerning the delimitation of LMAs (Casado-Díaz & Coombes, 2011). Empirically, LMAs are delineated based on commuting (travel-to-work) data between basic territorial units (BTUs; e.g., communes, municipalities and counties), which are often the smallest spatial unit at which data are available. Two objectives are generally sought in the delimitation of LMA. The first is *cohesion* or *inner integration* (Goodman, 1970): an LMA is composed of BTUs with strong commuting links. The second objective is

self-containment or *external perfection* (Goodman, 1970): an LMA should be comprised of BTUs that have weak or no commuting links with other LMAs. However, these two objectives are conflicting. Starting from a fully divided territory in which each BTU conforms to an LMA and following an aggregative procedure, as the LMA's size increases, self-containment levels can only increase; but from a certain point this also causes a progressive decrease in cohesion, as more BTUs with little interaction and larger travel distances are grouped together and the strength of the overall commuting interaction within the LMA becomes weaker. Finding an appropriate balance between self-containment and cohesion is therefore critical in the delineation of LMAs.

Accompanying these optimization objectives, other specific restrictions have to be met (EUROSTAT & Coombes, 1992). *Homogeneity* of LMAs ensures statistical comparability, and is typically achieved by establishing a *minimum population* or an *area size* requirement. Other three constraints are contiguity (i.e., LMAs must be spatially continuous), exhaustive territorial coverage (i.e., all BTUs must be part of an LMA) and absence of overlapping LMAs (i.e., each BTU must be part of only one LMA). Together, these constraints ensure and facilitate effective development, application, analysis and coordination of spatial policy-making.

While the delimitation of LMAs is a common practice in developed countries, lack of commuting data has prevented this tradition in most developing nations. Chile emerges as a unique country in the developing world. Unlike most developing countries, it has high-quality census data on commuting. Despite this, only two scholarly attempts have been made to create LMAs (Abalos & Paredes, 2012; Berdegúe et al., 2011), with both found to have major data and methodological shortcomings which are discussed below in the second section.

By redressing these deficiencies, this paper aims to build a set of LMAs for Chile using an optimization approach. Underpinning the significance of this work is a policy recommendation by the OECD for the Chilean government to focus on place-tailored initiatives (OECD, 2009). The proposed LMAs are a key step in this endeavour and provide local government agencies with a more effective spatial framework for policy development and monitoring of spatial labour market activity. The main novelty of this paper lies in the use and enhancement of an optimization procedure proposed by Martínez-Bernabeu, Flórez-Revuelta, and Casado-Díaz (2012) known as the grouping evolutionary algorithm (GEA).

The paper is structured as follows. The next section reviews the literature on LMA delimitation and highlights the advantages of evolutionary algorithms compared with commonly used methods. The third section presents the optimization problem (i.e., the delineation of an appropriate set of LMAs for a given territory) and describes GEA. This section also proposes two improvements to the GEA methodology: (1) a modification of the interaction index used in the problem's fitness function; and (2) the adoption of a spatially structured population model in the evolutionary algorithm. Together, these refinements improve the efficiency of GEA and the quality of LMA delimitations. The fourth section describes the data used. The fifth section provides evidence in support of our suggested improvements versus four competing model configurations and sets the model requirements of self-containment and population size. This section also presents our proposed 62 LMAs for Chile and assesses their performance against administrative areas (communes and provinces) and functional regions based on a commonly used algorithm, the travel-to-work areas (TTWAs) method. The sixth section provides some concluding remarks.

LITERATURE REVIEW

The delimitation of LMAs by public administrations has become a common practice in various developed countries, such as the United States (labour market areas), the UK (TTWAs), France (*zones d'emploi*), Germany (*arbeitsmarktreionen*), Italy (*sistemi locali del lavoro*) and Sweden (*lokala arbetsmarknader*). The OECD (2002) and Casado-Díaz and Coombes (2011) deliver exhaustive

reviews of the methodologies and criteria that have been employed by government agencies for the development of LMAs. Three key conclusions emerge from these reviews. First, LMAs represent a useful spatial framework for the design, implementation and monitoring of labour market policies and for statistical analysis of labour market activity at local levels. Second, despite a growing number of empirical applications, a lack of consensus is identified as to the most appropriate method to delineate LMAs. Different countries employ different methods, although it is remarkable that one specific method, the so-called TTWAs procedure, which has been used for the official delineation of LMAs in the UK (Coombes & Bond, 2008; Office for National Statistics (ONS), 2015), has recently become the official procedure also in Italy (Istituto Nazionale di Statistica (ISTAT), 2015). Third, while maximizing cohesion and self-containment are widely acknowledged to be the key principles for LMA delineation, existing procedures do not measure cohesion, and only some establish requirements of self-containment and/or size for the resulting set of LMAs. Even in attempts that quantify self-containment (Coombes & Bond, 2008; Coombes, Green, & Openshaw, 1986; ISTAT, 2005, 2015), variations in the required minimum levels are identified across countries and over time. Thus, there are not established standard values for self-containment and size.

The recent academic literature has focused on the development of new methodologies and application of existing methods to new datasets. Applications for a number of countries can be found: Australia (Watts, 2013), the Czech Republic (Klapka, Halás, Erlebach, Tonev, & Bednář, 2014), Estonia (Novak, Ahas, Aasa, & Silm, 2013), France (Fusco & Cagliani, 2011), Germany (Kropp & Schwengler, 2014), Greece (Prodromidis, 2008), Ireland (Farmer & Fotheringham, 2011), Slovenia (Drobne, Konjar, & Lisec, 2010), Spain (Feria-Toribio, Casado-Díaz, & Martínez-Bernabeu, 2015; Martínez-Bernabeu & Casado-Díaz, 2016) and the UK (Casado-Díaz, Martínez-Bernabeu, & Flórez-Revuelta, 2016), among others. Comparative analyses – both applications of alternative methodologies to one dataset or of one methodology to a range of datasets – are scarce. Thus, there is no agreement on a best approach for LMA delimitation.

Most LMA delimitation studies used methods based on the application of greedy decision rules. These rules force the amalgamation of the starting BTUs to form LMAs based on local maxima and mergers are determined without reconsideration of previous decisions. An example is the hierarchical aggregation procedure known as Intramax (Masser & Brown, 1975; see Landré, 2012, for an extensive list of applications). In this family of methods, mergers at each step are locally optimal – the pair of groups of BTUs that maximize a form of interaction index is merged. However, this does not guarantee a global optimum solution, as it is possible to perform adjustments to a resulting set of LMAs that improve the overall quality of the delimitation in terms of cohesion and self-containment (Watts, 2013).

Optimization algorithms for LMA delimitation

To achieve delimitations closer to the global optimum, there is a growing literature on the application of approximate optimization techniques to delimit LMAs drawing on operation research and artificial intelligence (Alonso, Beamonte, Gargallo, & Salvador, 2015; Chakraborty et al., 2013; Farmer & Fotheringham, 2011; Flórez-Revuelta, Casado-Díaz, & Martínez-Bernabeu, 2008; Fusco & Cagliani, 2011; Martínez-Bernabeu et al., 2012). These techniques search for an optimal LMA delimitation or its best approximation based on an objective or *fitness* function. This function yields a quantitative evaluation of a complete delimitation that can be used to compare and rank all valid alternative delimitations. Thus, these methods assess the quality of the entire set of LMAs, rather than exclusively relying on decisions based on local indicators.

Optimal LMA delimitation is a *partitioning problem*. Similar partitioning problems have been shown to be NP-complete, as is the case of the well-known *graph partitioning problem* (Fjallstrom, 1998). In this kind of combinatorial problems, the fitness function is to be maximized respect to the possible partitions of a set of elements (in our case, BTUs into LMAs). The computational

time required to solve this kind of problems through exact optimization makes its application infeasible when N is in the order of hundreds or greater. In a typical real case LMA delimitation, the number of BTUs to be partitioned is in the order of hundreds or thousands (Chile has 304 BTUs). Therefore, this problem must be tackled through approximate optimization techniques.

Fusco and Cagliani (2011) and Farmer and Fotheringham (2011) used the modularity quality function to guide different optimization techniques. Proposed by Newman and Girvan (2004), this function was developed in the frame of social networks for the detection of *communities* (groups of *nodes* with high interaction between themselves and low interaction with nodes in other communities). The modularity quality function accumulates the difference between the interaction links within each community and their expected value in a network with the same nodes with uniformly distributed flows (the null model). In the LMA context, nodes are conceived as BTUs and communities as LMAs. Fusco and Cagliani (2011) applied the optimization method proposed by Blondel, Guillaume, Lambiotte, and Lefebvre (2008). As is the case with Intramax, this method first considers each BTU as a potential LMA, and then iteratively aggregates the pair of potential LMAs that maximizes the increase in modularity. The process stops when no further improvement in the modularity quality function is recorded. Farmer and Fotheringham (2011) applied the method originally proposed by Newman and Girvan (2004), based on spectral bi-partitioning and recursive refinement of solutions. It begins by considering all BTUs as one single LMA, then recursively divides into two the LMA that maximizes the increase in modularity until no further divisions are possible. Following this process, the resulting hierarchical partitioning tree is examined backwards to identify and perform possible reallocations of single BTUs that increase the modularity score.

The use of the modularity function has been criticized in the context of community detection (Fortunato & Barthelemy, 2007; Lancichinetti & Fortunato, 2011). A major criticism is that modularity suffers from two coexisting problems: the tendency to merge small sub-graphs [LMAs], which dominates when the resolution is low; and the tendency to split large sub-graphs [LMAs], which dominates when the resolution is high (Lancichinetti & Fortunato, 2011). These problems prevent arriving at an optimal solution, even when actual communities are readily identifiable by alternative methods and visual inspection.

These issues are likely to be more acute in the context of LMA delimitation because the modularity function was developed for social networks that neglected geographical distances. In the null model used in the modularity function, the population size of a territory determines the expected sizes of commuting flows. In a sufficiently large territory, the expected commuting flows between pairs of BTUs or groups of BTUs in the null model would be very small, in a way that almost any existing commuting flow could force mergers (of actually separated LMAs) until only a few macro-regions cover the whole territory.

Flórez-Revuelta et al. (2008) made the first contribution to the set of optimizing LMA regionalization approaches by proposing an evolutionary algorithm drawn from artificial intelligence based on a new fitness function. Martínez-Bernabeu et al. (2012) refined this approach, developing a GEA. Using the same fitness function the method was then adapted by Chakraborty et al. (2013) and Alonso et al. (2015). The fitness function in these approaches operates accumulating the commuting interaction between each BTU and other BTUs within an LMA and thus provides a direct treatment of the cohesion objective in the optimization process. The interaction between BTUs is measured through the same index used in the TTWA method. This index measures the intensity of commuting flows between two groups of BTUs relative to the size of each group. Therefore, the fitness function based on this index is not affected by the problems associated to the modularity function.

Evolutionary algorithms are a general-purpose optimization technique inspired by the mechanisms of natural evolution. Starting from a set of initial solutions to the problem being solved (a *population* of *individuals* in the specialized terminology), new ones are produced iteratively by

recombination of previous solutions and the introduction of stochastic changes (*mutations*), and they all compete in terms of a given fitness function to be selected to remain for the next iteration. Thus, the solutions that remain gradually increase the average and maximum fitness scores, effectively performing an optimization process. The process finishes when a stop condition (e.g., elapsed time or stagnancy in the evolution) is met and the best solution so far is returned.

Prior LMA proposals for Chile

For Chile, two attempts have been made to define LMAs but they are hindered by methodological and data limitations. Using census data, Berdegué et al. (2011) delivered a first attempt using a hierarchical clustering based on Tolbert and Sizer's (1987) method. This method is similar to Intramax but based on an alternative interaction index which measures the commuting flow between two groups of BTUs divided by the working population of the smallest group. Abalos and Paredes (2012) delivered a second attempt using commuting data from the 2009 Chilean National Characterization Survey (CASEN) using an Intramax algorithm.

A major shortcoming of both attempts is that they applied methodologies based on non-optimizing procedures. As discussed above, these procedures are exclusively guided by local decisions and therefore they do not guarantee a global optimum. Moreover, they produced sets of LMAs that are likely to contain areas with poor levels of self-containment which conflicts with one of the main objectives of LMA delimitation. This is particularly evident by the high fragmentation of the Santiago Metropolitan Region (one of the 15 Chilean major administrative regions). Abalos and Paredes (2012) partitioned this area into five LMAs and Berdegué et al. (2011) divided this region into six non-contiguous LMAs. However, most of these communes constitute what it is commonly referred to as Greater Santiago, the metropolitan area around the capital city, and are usually grouped together for planning purposes due to their high commuting, transport and trade interaction, and geographical adjacency (Zegras & Gakenheimer, 2000).

A further shortcoming in Abalos and Paredes (2012) is the use of the CASEN survey data. This survey comprises a sample of only 1.5% of the Chilean population, and thus provides a highly sparse origin–destination matrix of commuting flows (Rowe, 2013). This has the potential to produce biased results due to statistical under-representation of small and medium-sized communes. Moreover, the final set of LMAs proposed by Abalos and Paredes (2012) is based on an origin–destination commuting matrix of only geographically contiguous communes. This is likely to produce spurious results as commuting flows between non-adjacent BTUs are also common and in some cases larger than those involving contiguous BTUs.

To overcome these deficiencies, we use census data and an enhanced variant of GEA developed by Martínez-Bernabeu et al. (2012). Unlike commonly applied methods based on local decisions, this approach is a global optimization method and utilizes a fitness function that is not affected by the limitations of the modularity fitness function.

METHODOLOGY

To define LMAs, we use and extend the GEA method developed by Martínez-Bernabeu et al. (2012) that builds on previous work by Flórez-Revuelta et al. (2008). This approach involves a maximization problem to group BTUs into LMAs subject to restrictions in terms of self-containment and population size via evolutionary computation techniques. The next subsection presents the optimization problem, its restrictions and fitness function. The following subsection describes the GEA optimization algorithm. The last subsection describes our proposal to enhance this methodology. First, we propose to improve the fitness function optimization problem by developing a new interaction index to measure within-LMAs cohesion. This index enables one to reach a better balance between the conflicting objectives of cohesion, self-containment and homogeneity.

Second, we propose to improve the performance of the evolutionary algorithm by adopting a spatially structured model as described below.

Optimization problem formulation

Let U be the set of BTUs that constitute a given territory. The general problem is to find the partition R of U into contiguous regions so that the accumulated interaction between each BTU and the rest of its LMA is maximal, subject to certain constraints for the LMAs: they must be continuous and reach a minimum size and self-containment. The input data are a square matrix T of order $|U|$, where U is the set of BTUs in the territory; and T_{ij} is the number of residents of BTU i that work in BTU j . With this notation,

$$T_{X,Y} = \sum_{i \in X} \sum_{j \in Y} T_{ij}$$

is the number of workers commuting from LMA X to LMA Y ;

$$O_X = \sum_{i \in X} \sum_{j \in U} T_{ij}$$

is the total number of workers living in X ; and

$$J_X = \sum_{i \in X} \sum_{j \in U} T_{ij}$$

is the total number of jobs in X .

This optimization problem can be formulated as follows:

Maximise $f(R)$

Subject to

$$O_A \geq t_{\min} \forall A \in R \quad (1)$$

$$\text{msc}(A) = \min\left(\frac{T_{A,A}}{O_A}; \frac{T_{A,A}}{J_A}\right) \geq \text{sc}_{\min} \forall A \in R \quad (2)$$

where A represents each of the LMAs in R ; and f is the fitness function that measures the accumulated interaction; and equations (1) and (2) define the numerical restrictions.

Constraints

In accordance with the existing literature (Coombes & Bond, 2008; Coombes et al., 1986; EUROSTAT & Coombes, 1992; ISTAT, 2005, 2015; OECD, 2002; ONS, 2015), equation (1) expresses the restriction of minimum number of occupied residents per LMA (where t_{\min} is the required value), and equation (2) the restriction of minimum self-containment per LMA (where sc_{\min} is the required minimum value). Self-containment of a region is defined as the minimum value of two metrics: supply-side self-containment – the number of people living and working in the region divided by the number of residents in the region (T_{AA}/O_A); and demand-side self-containment – the number of people living and working in the region divided by the number of jobs in the region (T_{AA}/J_A).

The three additional constraints typical of this problem are also applied: exhaustive coverage of the territory, absence of overlapping between LMAs and contiguity within LMAs. The first two are implicitly ensured as the implementation of GEA only allows a BTU to be in one and only one LMA. Contiguity is ensured throughout the optimization process described below by only permitting the reassignment of a given BTU from one LMA to another when it is adjacent to the ‘receiving’ LMA and the ‘sending’ LMA is still continuous after such reassignment.

Fitness functions

Two fitness functions originally proposed by Flórez-Revuelta et al. (2008) have been used in this optimization problem. The first function (f_1) is defined as the accumulated summation of an interaction index between each BTU and remaining BTUs within an LMA. It represents a clear innovation over previous methods, measuring not only self-containment but also cohesion within each LMA. This function¹ is expressed as:

$$f_1(R) = \frac{1}{U} \sum_{A \in R} \sum_{i \in A} I^{\text{Smart}}(i, A^{(i)}) \quad (3)$$

where i is each of the BTUs in A ; $A^{(i)}$ is the region composed by the BTUs of $A - i$; and I^{Smart} is an interaction index, a measure of commuting dependence between two areas relative to their size in terms of commuters. This index was originally proposed by Smart (1974) and subsequently used in the delimitation of British TTWAs. It is defined as:

$$I^{\text{Smart}}(A, B) I^{\text{Smart}}(B, A) = \frac{T_{A,B}^2}{O_A \cdot J_B} + \frac{T_{B,A}^2}{O_B \cdot J_A} \quad (4)$$

where the first term of the summation is the percentage of residents from A who work in B ($T_{A,B}/O_A$) times the percentage of jobs in B that are held by residents from A ($T_{A,B}/J_B$). The second term of the summation is the percentage of residents from B who work in A ($T_{B,A}/O_B$) times the percentage of jobs in A that are held by residents from B ($T_{B,A}/J_A$). This index allows one to measure the relevance of the bidirectional commuting flow between two regions in relation to the total number of commuters living or working in these regions.

The second fitness function (f_2) used by Flórez-Revuelta et al. (2008) is defined as f_1 times the total number of LMAs:

$$f_2(R) = \frac{\text{card}(R)}{U} \sum_{A \in R} \sum_{i \in A} I^{\text{Smart}}(i, A^{(i)}) \quad (5)$$

The term $\text{card}(R)$ represents the cardinality of a set of elements, specifically the number of LMAs in the regionalization R . Due to the inclusion of this term in f_2 , the use of this function leads to a larger number of LMAs compared with those resulting from the application of f_1 , at the expense of lower self-containment.

GEA optimization algorithm

As indicated in the second section, GEA tackles this problem through approximate optimization using evolutionary computation in a context where exact optimization procedures are not feasible due to the problem's complexity. Departing from a set of alternative LMA delimitations, GEA iteratively generates new delimitations through the application of sub-algorithms (called operators) that make stochastic changes, evaluates the new delimitations and selects those that will remain for the next iteration, favouring the delimitations with better fitness values. This process progressively improves the quality of the delimitations, and the procedure stops after a certain number of consecutive iterations are performed without an improvement in the best fitness value, which becomes the final delimitation.

GEA's operation can be better explained by breaking it into (1) the creation of an initial set of P delimitations by applying P times a stochastic hierarchical agglomerative (SHA) algorithm; (2) the main loop of the evolutionary algorithm that performs the optimization process; and (3) the set of operators (sub-algorithms) that generate the new s within the main loop. These three elements are described below.

SHA algorithm

- Create a new delimitation by considering each BTU as an independent LMA. Calculate the degree of validity (equation 6) for all the LMAs.
- While there is an LMA with validity < 1 , select at random an LMA A with validity < 1 and another adjacent LMA B with high² interaction (equation 4). Then merge A and B into a new LMA and calculate its degree of validity.

The degree of validity of an LMA is calculated as follows (Coombes & Bond, 2008):

$$\text{validity}(A) = \min\left(\frac{\text{msc}(A)}{sc_{\min}}; 1\right) \times \min\left(\frac{O_A}{t_{\min}}; 1\right) \quad (6)$$

Main loop

After the creation of the initial delimitations, GEA iterates over the following steps until the last G iterations do not produce an improvement of the best delimitations:

- From the pool of P delimitations, select one with probability proportional to its fitness value and select at random one of the 10 operators (described below). If the operator is *recombination* then select at random another delimitation.
- Apply the operator to generate a new delimitation, calculate the fitness value of the new delimitation and add it to the current pool of P delimitations.
- From the current $P + 1$ delimitations, select the one with highest fitness value and another $P - 1$ with probability proportional to their fitness values to form the pool of P delimitations that remain for the next iteration.

Operators

The 10 operators that GEA use to create new delimitations in its main loop (Martínez-Bernabeu et al., 2012) can be summarized as follows:

- Partition: (i) select a random LMA A; (ii) select at random another LMA adjacent to A with non-null interaction; and (iii) delete both LMAs and apply SHA algorithm to the BTUs from the deleted LMAs to get a new grouping of them.
- New region: (i) select at random a border BTU (a *border BTU* is a BTU within an LMA that is adjacent to BTUs in a different LMA); (ii) create a new LMA A with that BTU; and (iii) try to reassign into A adjacent BTUs with non-null interaction, randomly selected from neighbouring LMAs, until the new LMA is valid (success) or there are no more adjacent BTUs that can be reassigned without breaking the validity of any LMA (fail, no new regionalization is produced).
- Dismember: (i) select a random LMA A; and (ii) while there are BTUs in A, select at random one of its border BTUs and reassign it to an adjacent LMA with high interaction.
- Exclusion: (i) select a random LMA A; (ii) select at random one of its border BTUs that could be reallocated to an adjacent LMA without breaking validity; and (iii) if this is not possible then finish; otherwise perform the reallocation and go back to (ii).
- Inclusion: (i) select a random LMA A; (ii) select at random one adjacent BTU from the adjacent LMAs that could be reallocated to A without breaking validity; and (iii) if this is not possible then finish; otherwise perform the reallocation and go back to (ii).
- Segregation: (i) select a random LMA A with high population and an adjacent LMA B with high interaction; (ii) select at random one BTU from A adjacent to B that could be reallocated

- without breaking validity; and (iii) if this is not possible then finish; otherwise perform the reallocation and go back to (ii).
- Annexation: (i) select a random LMA A with low self-containment and an adjacent LMA B with high interaction; (ii) select at random one BTU from B adjacent to A that could be reallocated without breaking validity; and (iii) if this is not possible then finish; otherwise perform the reallocation and go back to (ii).
 - Mutation: (i) select a random number r between 1 and $U/50$; and (ii) repeat r times: select at random a border BTU and reallocate it to an adjacent LMA with high interaction if that does not break validity.
 - Exchange: (i) select one LMA A with low self-containment and an adjacent LMA B with high interaction; (ii) select one BTU from A adjacent to B and reallocate it (even if it breaks validity); (iii) select a different BTU from B adjacent to A and reallocate it; and (iv) finish successfully if both LMAs remain valid; otherwise no new regionalization is produced.
 - Recombination (the only operator that works over two different regionalizations): (i) create a new regionalization as an exact copy of one of the regionalizations; (ii) select from the other regionalization a random number of LMAs between 1 (absolute number) and 66% of its LMAs and copy them into the new regionalization – this usually breaks the validity of some of the initial LMAs in the new regionalization because portions of them have been reallocated; and (iii) the SHA algorithm is applied to repair invalid LMA (fragments of previously valid LMAs) by merging them with adjacent LMAs until they all are valid.

Improvements to GEA methodology

In this paper we propose two improvements to the approach developed by Martínez-Bernabeu et al. (2012) to increase the computational efficiency of the evolutionary algorithm and to improve the quality of the final regionalizations in terms of the number of misallocated BTUs and the balance between the different objectives, as explained below.

A new fitness function

Our first, methodological improvement is a modification to the fitness function originally proposed by Flórez-Revuelta et al. (2008). A drawback of both functions f_1 and f_2 is their tendency towards the production of a number of misallocated BTUs (this effect is illustrated below in the fifth section using Chilean data). A BTU is considered to be misallocated when its commuting interaction with a neighbouring LMA exceeds its interaction with the rest of its LMA. These BTUs could therefore be allocated to a more suitable LMA in such a way that would improve their local self-containment and cohesion statistics. However, such reassignment would lead to a slight deterioration in the fitness function. This is because although the interaction scores of reassigned BTUs would improve, it would simultaneously slightly worsen the interaction indices of other BTUs, offsetting the resulting improvement. On balance, because of this undesirable trade-off, the optimization process based on f_1 and f_2 favours the regionalization with the misallocated BTUs. However, we argue that the significant improvement in the interaction index of the reassigned BTUs is worth the small drops in the interaction indices of the other affected BTUs.

Thus, we propose a different form of the interaction index departing from the Smart's interaction index (equation 4). As explained above, this index is the sum of two products of proportions; that is, the proportion of residents in A working in B times the proportion of jobs in B being held by workers from A, plus the proportion of residents in B working in A times the proportion of jobs in A being held by workers from B. To facilitate interpretation, we express this index as a proportion by dividing by two and applying the square root (which provides a

measure analogous to the geometric average):

$$I^{\text{SRS}}(A, B)I^{\text{SRS}}(B, A) = \sqrt{\frac{I^{\text{Smart}}(A, B)}{2}} \quad (7)$$

This allows a direct comparison between commuting percentages and interaction values. Using this index (equation 7) within f_1 we obtain a new fitness function f_3 :

$$f_3(R) = \frac{1}{U} \sum_{A \in R} \sum_{i \in A} I^{\text{SRS}}(i, A^{(i)}) \quad (8)$$

The experimentation conducted for the Chilean case (see below) revealed that the application of this fitness function leads to increases in LMAs' homogeneity and self-containment at the expense of a small reduction in the number of identified LMAs. It also showed reductions in the number of misallocated BTUs by mitigating the effects of undesired trade-offs between cohesion and self-containment.

For our comparative assessment of fitness functions, fully developed below, we derived a fitness function f_4 from f_3 in the same way f_2 was derived from f_1 ; that is, f_4 is f_3 times the total number of identified LMAs:

$$f_4(R) = \frac{\text{card}(R)}{U} \sum_{A \in R} \sum_{i \in A} I^{\text{SRS}}(i, A) \quad (9)$$

A spatially structured model

Our second enhancement of the approach involves the adoption of a parallel, spatially structured model (Tomassini, 2005). This change affects the way in which regionalizations are chosen to produce new ones by recombination and to remain for the next iteration. This model is recognized to significantly improve the performance of evolutionary algorithms by increasing the diversity of solutions (Alba, 2002). While the evolutionary approach used by Martínez-Bernabeu et al. (2012) is based on a *panmictic* model (there is only a single set of regionalizations in which all P regionalizations compete with each other to be selected for the next iteration), in a spatially structured model regionalizations are divided into several subsets, each of which operates relatively independent from the rest. The main loop of GEA is applied independently and in parallel to each subset of regionalizations, so that regionalizations on each subset only compete with others in the same subset. An extra step is added to the main loop of GEA, in which the recombination is applied to regionalizations belonging to different subsets with a much lower probability than the regular recombination within subsets. This grants more time for suboptimal delimitations on each subset to evolve and surpass other local maxima in different subsets before being excluded from the evolution. As a result, there is an increase in the diversity of delimitations that helps the algorithm to not converge to local maxima but find a global optimum, improving the efficiency and efficacy of the evolutionary algorithms, as shown in the fifth section.

To implement our spatially structured model, subsets of LMA delimitations are organized in a parameterized grid. This defines a topological space comprising a set of neighbourhoods for each subset: each subset has another four neighbouring subsets: up, down, left and right. In the step added to the main loop, when a recombination between subsets is triggered for a certain subset, a neighbouring subset is randomly uniformly selected. Then, one LMA delimitation of each subset is selected for the *recombination* to create a new one, that is placed in the subset that triggered this recombination.

Table 1 lists the parameters that need to be specified to control the optimization process of GEA.

Table 1. Grouping evolutionary algorithm (GEA) configuration parameters.

| Parameter | Description |
|------------------|---|
| $D_1 \times D_2$ | Dimensions of a two-dimensional grid of subsets of regionalizations |
| M | Frequency of application of recombination between subsets |
| P | Number of regionalizations in each subset |
| N | Number of operations (new regionalizations) per subset and iteration |
| G | Maximum number of iterations without improvement in the best solution to stop |

Note: The panmictic model can be seen as a special case of the spatially structured model in which parameters D_1 and D_2 are equal to 1, and M is not used. The total number of regionalizations across all subsets is $D_1 \times D_2 \times P$.

DATA

To delineate LMAs for Chile, we draw on commuting data from the Chilean Migration (CHIM) database (Rowe & Bell, 2013). The CHIM database provides data at three levels of spatial scale: 304 communes, 51 provinces and 13 regions.³ The commuting data are only available for 2002:⁴ this was the first time information on respondents' place of residence and employment was simultaneously recorded in the Chilean census. No data on frequency, duration and mode of commuting transport was recorded, thus our data include travel-to-work trips of different length of stay, distance and timing. However, 90% of these commuting flows covered a distance of less than 60 km (Rowe, 2014; Rowe & Bell, 2017), suggesting that most of these moves comprise a daily commuting pattern. For our analysis, we used commuting data at the lowest administrative geography available: communes (our BTUs). From the 304 communes in the data, 301 communes were used to produce an origin–destination matrix of commuting flows.⁵

To be able to impose the contiguity constraint, we created a binary spatial contiguity matrix to identify geographically adjacent BTUs. In a few cases, strict spatial contiguity was not present due to separating water channels. Time distances by road were not used because they do not reflect the geographical accessibility of BTUs in southern Chile. Many of these BTUs have no road connection but are linked to other BTUs through ferry lines. Thus, spatial adjacency was manually specified for those cases connections through ship transport (Porvenir and Punta Arenas, Dalcahue and Curaco de Vélez, Quellón and Guaitecas, and Chonchi and Puqueldón). Establishing spatial contiguity in this way provides a better representation of the real commuting linkages between BTUs with no road connection.

COMPARATIVE ASSESSMENTS AND OUTCOMES

This section first performs assessments of our enhancements to the evolutionary algorithm and fitness function, providing evidence of their significance in the first two subsections. Then it assesses and discusses the selection of appropriate thresholds for self-containment and population size to define LMAs for Chile in the third subsection. Finally, the fourth subsection presents our final proposal of LMAs, which is evaluated against administrative areas (BTUs and provinces) and the LMA regionalization produced by the most widely used method for LMA delimitation, the TTWAs method.

To provide a quantitative foundation for our assessments, we use a range of global indicators that capture the three key dimensions of self-containment, cohesion and homogeneity across the entire system of LMAs. Table 2 lists and describes these indicators.

Panmictic versus spatially structured model

To assess improvements of the spatially structured model in the optimization process, we conducted a performance analysis. We compared the GEA based on spatially structured model versus

Table 2. Statistical measures of self-containment, cohesion and homogeneity.

| Objective | Indicator | Description and notes |
|----------------------------|--|--|
| Self-containment | Global self-containment $\sum_{A \in R} T_{A,A} / \sum_{A \in R} O_A$ | Proportion of people in the regionalization who work in their LMA of residence. Higher values are preferred as it implies a reduction in spillover |
| | Minimum and median self-containment across LMAs | Self-containment of an LMA is the proportion of people residing in that LMA who work within its borders. We use minimum and median values of this measure across the LMAs of a regionalization as an indicator of local self-containment |
| Cohesion | Global interaction index | Average of the interaction index between each BTU and the rest of its LMA (different versions depending on the interaction index used, see equations 1 and 5). Higher values imply higher interaction between the BTUs of each LMA |
| Workforce size homogeneity | Minimum, first decile, median, ninth decile and maximum number of occupied residents across LMAs | Homogeneity in workforce size is argued to improve comparability between LMAs |
| Area size homogeneity | Minimum, first decile, median, ninth decile and maximum geographical area (km ²) across LMAs | Too large LMAs imply long commuting distances and possibly low cohesion. Clusters of too small LMAs imply that mergers between nearby LMAs that could improve self-containment without deteriorating cohesion have not been performed |

the panmictic model using six experiments with different model parameter configurations, as displayed in Table 3. Configurations A–D are spatially structured, while E and F are panmictic. E is a *conservative* configuration with many alternative regionalizations (600), operations per iteration (200) and iterations without improvement to stop (4000). We label this configuration as conservative because it devotes a great computational effort to the optimization task. In contrast, F is labelled a *fast* configuration with fewer regionalizations (150), operations per iteration (50) and iterations to stop (1000). It requires less time to return a delimitation, although such delimitation could potentially be of less quality than that returned by the conservative configuration. Comparable scenarios were created for the spatially structured model, producing the same number of new regionalizations per iteration and maximum number of iterations without improvement. Configuration A matches E, and D matches F. Configurations B and C can be seen as intermediate experiments between A and D. B shares the conservative (bigger) grid of A but applies a fast stopping condition as D. C uses the same fast (smaller) grid as D but with A's conservative stopping condition.

We run each configuration 10 times using data for Chile, fitness function f_1 , a standard minimum self-containment threshold of 75% (Smart, 1974; ISTAT, 2005, 2015) and no size

Table 3. Comparative GEA analysis: a spatially structured population model versus a panmictic model.

| Configuration | Spatially structured population model | | | | Panmictic model | |
|-------------------------------------|---------------------------------------|----------------|---------------|---------------|-----------------|--------|
| | A | B | C | D | E | F |
| $D_1 \times D_2$ | 2×100 | 2×100 | 1×50 | 1×50 | 1 | 1 |
| P | 3 | 3 | 3 | 3 | 600 | 150 |
| Total regionalizations | 600 | 600 | 150 | 150 | 600 | 150 |
| Operations per subset and iteration | 1 | 1 | 1 | 1 | 200 | 50 |
| Total operations per iteration | 200 | 200 | 50 | 50 | 200 | 50 |
| Iterations without improvement | 4000 | 1000 | 4000 | 1000 | 4000 | 1000 |
| Resulting best fitness | 0.0436 | 0.0436 | 0.0436 | 0.0436 | 0.0436 | 0.0436 |
| Frequency of best solution | 1.0 | 0.8 | 1.0 | 0.4 | 0.9 | 0.5 |
| Average CPU time (s) | 205.3 | 113.3 | 45.2 | 24.4 | 68.6 | 7.0 |

Note: $sc_{\min} = 75\%$, $t_{\min} = 0$ and f_1 is used for all model configurations. Regionalizations per subset refers to the number of solutions in each subset. Operations per subset and iteration refers to the number of new regionalizations created on each subset per iteration.

Source: Authors' elaboration using 2002 census data.

requirement. As could be expected, the results show that by allowing sufficient computational time, any configuration can find the best delimitation, but particular configurations are shown to require more time. Thus, for example, when GEA configurations with 4000 iterations are compared, both spatially structured and panmictic models find the best delimitation in all or most repetitions. However, the spatially structured model outperforms the panmictic model in finding the best delimitation with a higher frequency. In terms of both efficiency and efficacy, the best configuration is C. It produced the lowest average CPU time across all the configurations that always find the best delimitation. Nevertheless, computational time for Chile is affordable for all model configurations given its small number of BTUs. Therefore, efficiency does not represent a critical decision factor in this instance. Rather, we decided to use a conservative model configuration (A) for all the remaining experimentation to ensure we obtain the best possible delimitations.

It must be noted that the GEA configuration parameters do not determine the characteristics of the final delimitation when sufficiently conservative values are used and only regulate the computational time required by the algorithm to return the final outcome, which is always the optimum (Table 3). In contrast, any change in the minimum thresholds of self-containment or size always has an effect on the final outcome, as will be shown below. A change in these thresholds implies building LMAs with different characteristics, with a different trade-off between the main objectives of LMA delimitation.

Fitness functions assessment

To evaluate our proposed fitness functions (f_3 and f_4), we performed a comparative assessment against f_1 and f_2 . Regionalizations based on these fitness functions were compared in terms of commuting self-containment, cohesion and homogeneity. Table 4 shows these statistics. A standard value of 75% for the minimum self-containment threshold, with no population size, was used to derive these regionalizations.

As expected, functions f_2 and f_4 , which use the number of identified LMAs as a factor, produce regionalizations that sacrifice self-containment and interaction in exchange for the identification of a greater number of LMAs. This causes a bias towards the creation of fragmented LMAs with

Table 4. Self-containment, cohesion and homogeneity measures of performance for four alternative fitness functions.

| Fitness function | | f_1 | f_2 | f_3 | f_4 |
|---------------------------------------|-------------|-----------|-----------|-----------|-----------|
| Identified regions | | 87 | 190 | 75 | 180 |
| Global self-containment (%) | | 94.25 | 93.11 | 94.65 | 93.31 |
| Global interaction index (equation 1) | | 0.00787 | 0.00718 | 0.00778 | 0.00740 |
| Global interaction index (equation 5) | | 0.04508 | 0.03491 | 0.04555 | 0.03769 |
| Misallocated BTUs | 27 | 25 | 16 | 23 | |
| Minimum self-containment by region | Minimum (%) | 75.69 | 75.00 | 75.48 | 75.00 |
| | Medium (%) | 89.60 | 85.65 | 90.35 | 86.34 |
| Area by region (km ²) | Minimum | 301 | 96 | 329 | 96 |
| | Decile 1 | 820 | 368 | 915 | 402 |
| | Medium | 2483 | 1288 | 3058 | 1424 |
| | Decile 9 | 22,829 | 8761 | 27,540 | 9170 |
| | Maximum | 94,772 | 56,617 | 94,772 | 56,617 |
| Occupied residents by region | Minimum | 424 | 146 | 567 | 146 |
| | Decile 1 | 2061 | 1092 | 4965 | 1092 |
| | Medium | 12,089 | 4578 | 17,810 | 4768 |
| | Decile 9 | 85,495 | 41,574 | 98,118 | 42,445 |
| | Maximum | 1,865,101 | 1,865,101 | 1,875,595 | 1,871,357 |

Note: $sc_{\min} = 75\%$ and $t_{\min} = 0$ used for all four cases.

Source: Authors' elaboration using 2002 census data.

no improvements in the desired levels of self-containment, cohesion and homogeneity, except for a greater number of LMAs. Consequently, we disregard these formulations.

Compared with f_1 , f_3 produced a smaller number of LMAs with slightly higher self-containment, similar cohesion levels⁶ and slightly better homogeneity. This improvement in homogeneity is reflected in larger median scores of occupied residents and area by region as a result of the amalgamation of small LMAs. But, the most remarkable difference is the reduction in the number of misallocated BTUs: only 16 for the f_3 -based regionalization against more than 20 for the other three fitness functions. The same comparative analysis of the four fitness functions using 80% and 85% minimum self-containment produced similar conclusions.⁷ We thus selected the proposed fitness function f_3 to guide GEA.

Establishing minimum self-containment and size thresholds

Establishing minimum thresholds of LMA self-containment and population size is challenging. Large enough values of these thresholds lead to large and less cohesive LMAs, while too small values enable the identification of too small or insufficiently self-contained LMAs. Underpinning this is the commuting interaction between regions which tends to become weaker as their size and distance travel within their boundaries rise. For instance, if all BTUs are clustered into a single LMA, self-containment is maximal (i.e., no commuting flows between LMAs) but cohesion is minimal (i.e., no commuting interaction between many of the BTUs within that LMA). In contrast, if each BTU constitutes one LMA, cohesion would be high (we always assume high interaction within a BTU) but self-containment would be low (i.e., high levels of commuting between LMAs). The challenge is thus to find a balance between these conflicting aims. In practice, self-

containment and size requirements are determined based on expert knowledge and qualitative judgement that may be subject to arbitrariness, and objective measures of cohesion are neglected.

Our strategy was to first establish an appropriate self-containment threshold by testing four alternatives with no population size restrictions, and then decide a minimum population size threshold. Table 5 shows results for three different self-containment alternatives.

Self-containment threshold

We first used the standard 75% threshold for self-containment, producing 75 LMAs. The global self-containment statistic for this regionalization is 94.65%, which is considerably high compared with European countries, such as the UK and Spain. This result could be expected given the large areas and lower population densities of Chilean BTUs, particularly in the extreme northern and southern areas of the country.

The resulting LMAs show large variations in population and area size, counteracting our objective of homogeneity and reflecting the large differences in population and geographical areas across BTUs. Northern and southern BTUs are sparsely populated, have small populations, large areas (some of over 20,000 km², running against the possibility of daily commuting trips) and high self-containment levels. The 75% self-containment threshold also caused many of the densely populated BTUs in central Chile to form LMAs with considerably lower self-containment levels than northern and southern LMAs and arguably too small land areas. These small LMAs would entail commuting trips within a distance of under 50 km and probably less than 25 km as commuting trips usually concentrate towards a single core BTU. Moreover, these small LMAs present relatively high levels of interaction with neighbouring LMAs of similar size, which justifies their amalgamation into more cohesive LMAs.

Table 5. Self-containment, cohesion and homogeneity measures for three alternative minimum self-containment requirements: 75%, 80% and 85%.

| Minimum self-containment requirements | | 75% | 80% | 85% |
|---------------------------------------|-------------|-----------|-----------|-----------|
| Identified regions | | 75 | 70 | 62 |
| Global self-containment (%) | | 94.65 | 95.12 | 95.92 |
| Global interaction index (equation 1) | | 0.00778 | 0.00755 | 0.00729 |
| Global interaction index (equation 5) | | 0.04555 | 0.04455 | 0.04363 |
| Misallocated BTUs | | 16 | 18 | 12 |
| Minimum self-containment by region | Minimum (%) | 75.48 | 80.37 | 85.22 |
| | Medium (%) | 90.35 | 90.79 | 91.55 |
| Area by region (km ²) | Minimum | 329 | 573 | 1113 |
| | Decile 1 | 915 | 1236 | 1640 |
| | Medium | 3058 | 3088 | 3838 |
| | Decile 9 | 27,540 | 28,670 | 30,187 |
| | Maximum | 94,772 | 94,772 | 94,772 |
| Occupied residents by region | Minimum | 567 | 567 | 567 |
| | Decile 1 | 4965 | 4807 | 6507 |
| | Medium | 17,810 | 18,893 | 19,318 |
| | Decile 9 | 98,118 | 99,130 | 100,684 |
| | Maximum | 1,875,595 | 1,901,882 | 1,945,072 |

Note: $t_{\min} = 0$ and f_3 used for all three cases.

Source: Authors' elaboration using 2002 census data.

Seeking to find a better trade-off between self-containment, cohesion and homogeneity, we tested self-containment thresholds of 80%, 85% and 90%. The 90% threshold was discarded because it produced excessively large LMAs with a single macro-LMA in northern Chile, covering above 20% of the national territory and grouping communes separated by more than 600 km. Such a large area cannot comprise a single labour market capturing daily commuting. This indicates that requiring 90% minimum self-containment is clearly excessive to ensure an adequate level of cohesion within LMAs.

The 80% and 85% self-containment thresholds produced better and relatively similar results (Table 5). The 85% threshold produced 62 LMAs and the 80% value produced 70 LMAs, reporting global self-containment statistics of 95.92% and 95.12%, respectively. Their main differences are in greater cohesion and lower self-containment for the 80% delimitation, but these differences are too small to decide for one self-containment threshold. Thus, we assessed their associated homogeneity. The results showed that the largest LMAs are similar for both thresholds, but minimum values in population and area size are larger for the 85% regionalization. This indicates that the 85% threshold produced not only an improvement in self-containment but also in homogeneity, without compromising the principle of cohesion. Moreover, the 85% threshold also showed a lower number of misallocated BTUs. We thus defined 85% as the appropriate minimum self-containment threshold for our delimitation of LMAs for Chile.

Population size

To establish our minimum population size requirement, we based our decision on research aiming to determine the appropriate minimum population size for Chilean communes in terms of efficient local administrative costs. Horst (2005) showed that communes with fewer than 20,000 inhabitants tend to be bureaucratically less efficient as they tend to report higher administration costs per inhabitant relative to communes with populations over 20,000. Examining 2002 census data, this population number is roughly equivalent to 5000 employed residents. We thus established 5000 workers as our minimum population size threshold.

LMAs for Chile: evaluation of alternatives

This section presents our final LMA delimitation of 62 GEA-LMAs. This outcome was derived⁸ from our proposed GEA method using a spatially structured population model and our new fitness function (f_3), based on 85% minimum self-containment and 5000 minimum employed residents. Figure 1 displays our proposed GEA-LMAs along with 50 provinces and 72 LMAs produced by the TTWA method (TTWA-LMAs) using the minimum thresholds applied in GEA. Table 6 reports the associated self-containment, cohesion and homogeneity statistics of these areas, including our 301 BTUs.

Overall, the results indicate that GEA-based LMAs represent a more appropriate geographical framework to capture spatial labour market interactions than BTUs, provinces and TTWA-LMAs. They also suggest that TTWA-LMAs provide a more suitable framework than communes and provinces. Communes have excessively low levels of self-containment, reflecting that they tend to be too small to represent LMAs because a large share of their residents work in different BTUs.

Provinces appear better suited than communes to represent LMAs. However, their levels of self-containment and commuting interaction are lower than those displayed by GEA-LMAs and TTWA-LMAs, and show a greater variation in workforce and area size. In fact, both GEA-LMAs and TTWA-LMAs exhibit high local values of self-containment, meeting our requirement of 85% minimum self-containment, while identifying a higher number of areas⁹ than provinces. Thus, both GEA-LMAs and TTWA-LMAs appear as better frameworks to capture spatial labour market activity in Chile than available administrative units.

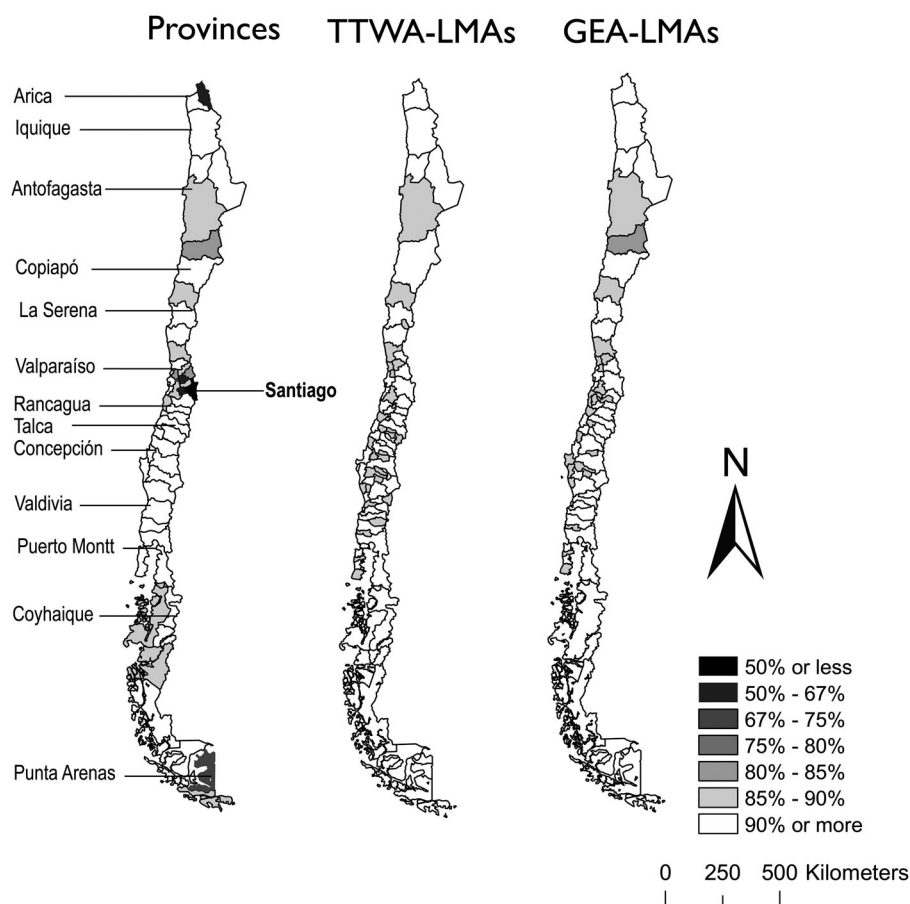


Figure 1. Maps of local self-containment: provinces, TTWA-LMAs and GEA-LMAs (labels for regional capital cities are displayed).

Note: Local self-containment of an area is calculated as the minimum value of the share of residents of a BTU working within its considered region, and the share of jobs of that BTU occupied by its residents.

When GEA-LMAs and TTWA-LMAs are compared, the results reported in Table 6 show that the former outperforms the latter, except for the number of identified LMAs. The TTWA method produced a greater number of LMAs than the GEA. This can be seen as a desirable outcome; however, this is overcompensated by considerably lower homogeneity in population and area size. The greater performance of GEA-LMAs becomes apparent through the analysis of local statistics of self-containment, cohesion and homogeneity. Compared with GEA-LMAs, TTWA-LMAs are formed by BTUs that have a higher degree of commuting interaction with BTUs outside their assigned LMA. For instance, Figure 2 shows that three TTWA-LMAs are produced in the Valparaíso region: Valparaíso, Quillota and San Antonio. These areas, however, report a high degree of commuting interaction between them, particularly in the resulting TTWA-LMAs of Valparaíso and Quillota. A share of up to 24% of residents commutes from the Quillota TTWA-LMA to the Valparaíso TTWA-LMA. In contrast, these BTUs are clustered to form a single LMA through GEA. While provinces also group these BTUs into a single LMA, they exclude other BTUs (Limache and Olmue) that display a considerably high commuting interaction with the Valparaíso province (out-commuting rates of 25% and 39% respectively).

Table 6. Self-containment, cohesion and homogeneity measures: BTUs, provinces, TTWA-LMAs and GEA-LMAs.

| Delimitation | | BTUs ^a | Provinces ^a | TTWA-LMAs ^c | GEA-LMAs ^c |
|--|-------------|-------------------|------------------------|------------------------|-----------------------|
| Identified regions | | 301 | 50 | 72 | 62 |
| Global self-containment (%) | | 60.94 | 90.21 | 95.75 | 95.88 |
| Global interaction index (equation 1) | | n.a. ^b | 0.00676 | 0.00700 | 0.00733 |
| Global interaction index (equation 5) | | n.a. ^b | 0.04066 | 0.04202 | 0.04555 |
| Minimum self-containment by region | Minimum (%) | 10.51 | 29.01 | 85.30 | 84.71 |
| | Median (%) | 79.79 | 91.51 | 90.95 | 91.31 |
| Regions with minimum self-containment <85% (%) | | 67.77 | 18.00 | 0.00 | 0.00 |
| Regions with minimum self-containment <70% (%) | | 21.56 | 12.00 | 0.00 | 0.00 |
| Area by region (km ²) | Minimum | 6 | 580 | 553 | 906 |
| | Median | 825 | 10,110 | 1454 | 4026 |
| | Maximum | 50,876 | 67,683 | 76,152 | 94,772 |
| Occupied residents by region | Minimum | 146 | 663 | 5127 | 5100 |
| | Median | 4512 | 44,312 | 6540 | 23,319 |
| | Maximum | 181,820 | 1,556,008 | 1,952,138 | 1,952,138 |

Notes: ^aCommunes and provinces as in CHIM database, excluding three non-continental communes: Isla de Pascua, Juan Fernández and Antártica.

^bThe inner interaction index cannot be calculated for regions composed of only one BTU.

^cTTWA and GEA LMAs were derived using $sc_{min} = 85\%$ and $t_{min} = 5000$. GEA used fitness function f_3 .

Source: Authors' elaboration using 2002 census data.

Additionally, Figure 2 shows that provinces in central Chile are too small and 'open' to represent spatial labour markets, particularly within the Santiago Metropolitan Region. Provinces in the Santiago Metropolitan Region display a low degree of self-containment, with more than 30% of workers commuting to other provinces and some even more than 70%. On the other hand, provinces in the region of Los Lagos and Los Rios in central-southern Chile appear to be too large, comprising only five areas and clustering two regional cities, Valdivia and Puerto Montt. These provinces are likely to be internally segmented, and their subdivision seems to provide a more appropriate geographical framework. The geographical extent of their associated commuting flows is highly localized and thus these five provinces are not sufficiently cohesive to represent separate LMAs.

Figure 2 also reveals that GEA-LMAs more accurately reflect the intensive labour market interaction between BTUs within the Metropolitan Region in central Chile by forming a small peripheral LMA of four BTUs and a large single LMA around Santiago through the amalgamation of 44 BTUs. Consistently, this last set of areas includes what is typically referred as to the Greater Santiago by local planning authorities to capture the transport linkages and spatial labour market interaction in this region (Zegras & Gakenheimer, 2000). We note that a similar LMA around Santiago is created using the various self-containment restrictions reported below in the fifth section: ranging from 36 BTUs, 5,647,590 inhabitants and 9270 km² for a 75% self-containment requirement, to 43 BTUs, 5,904,579 inhabitants and 11,321 km² for a 85% requirement, which always include the metropolitan area of Greater Santiago.

Conciliating all the conflicting objectives and restrictions in LMA delimitation is, however, a challenging task. We identified six excessively large GEA-LMAs (Iquique, Calama, Antofagasta, Coyhaique, Natales and Punta Areas). Each of these LMAs covers areas of at least 25,000 km, involving commutes of 250 km, and could be considered for subdivision. This action is, however,

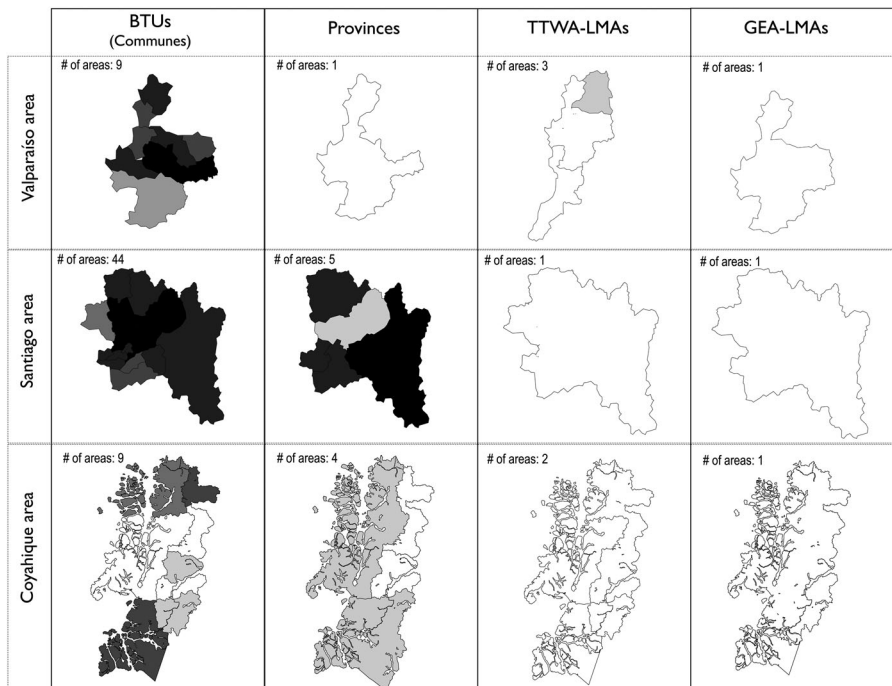


Figure 2. Maps of selected areas: number of regions and local self-containment measures. BTUs, provinces, TTWA-LMAs and GEA-LMAs.

reasonable for only the two biggest GEA-LMAs (Coyhaique and Natales), if considered necessary for policy-making or statistical purposes. The remaining four GEA-LMAs have one main BTU that concentrates commuting inflows within a daily commutable distance (<150 km) from all other BTUs within the same LMA. In contrast, the GEA-LMAs of Coyhaique and Natales were comprised of BTUs amalgamated in order to comply with our population restrictions, but the commuting links between some of their constituent BTUs are weak. By analysing¹⁰ commuting flows and distances between the constituent BTUs and their self-containment statistics, a potential subdivision could be proposed for the Coyhaique GEA-LMA by amalgamating the BTUs of Aisén and Coyhaique (which display high commuting interaction and high geographical accessibility from each other; one hour by road) and conserving the other BTUs (with small resident population but considerably self-contained and distant from the rest) as special cases of independent LMAs. For the Natales GEA-LMA, a subdivision would be considered by clustering the BTUs of Natales and Torres del Paine, and retaining the BTU of O'Higgins as another special case of independent LMA. We note that these subdivisions would imply violating the imposed requirements of minimum population size. We therefore do not implement these subdivisions for our proposal of 62 GEA-LMAs.

Taken together, our results indicate that the GEA-based LMAs provide a more appropriate spatial framework to analyse labour market activity in Chile than communes, provinces and TTWA-LMAs.

CONCLUSIONS

In 2009, the OECD assessed of Chile's territorial economic performance, revealing underutilized potential of regional assists to boost economic productivity (OECD, 2009). To improve public

management of these resources, the OECD recommends moving towards a territorial policy approach that capitalizes on the opportunities and needs of the country's diverse geography. Gaining a better understanding of the underlying spatial processes shaping Chile's economic geography and building adequate analytical tools are conceived as two critical steps to crystallize this change in policy approach. This paper aimed to make a contribution to these steps by developing a set of functional LMAs.

We contributed to the literature of LMA regionalization by applying and enhancing an evolutionary computational approach known as GEA. Compared with commonly applied methodologies that are based on heuristic processes, this evolutionary approach optimizes the global quality of the resulting regionalization through the maximization of a fitness function that explicitly integrates the main objective of cohesion.

We presented evidence supporting that our proposed LMAs provide a more appropriate framework for the study of labour market interactions than administrative units, the regionalization produced by the most common functional procedure officially in use (the TTWA method), and previous attempts to build functional regions for Chile.

Additionally, we made two methodological improvements to the original approach via (1) the reformulation of the regionalization optimization problem through the development of a new interaction index to build an enhanced fitness function, and (2) the adoption of a spatially structured model in the evolutionary algorithm. The first improvement allows a better conciliation between the different objectives of regionalization, and improves the local quality of the final regionalization by reducing the number of misallocated BTUs and the need to make adjustments to achieve a definitive regionalization. The second improvement increases the efficiency of the evolutionary algorithm by reducing the computational time required to find the optimal regionalization. Despite the novelty of this methodology in this field, improvements in the quality of regionalization outcomes are likely to encourage its application among LMA delimitation practitioners.

Building on the LMAs developed in this paper, the ultimate aim of our research programme is to trace the evolution of the interaction of labour demand and supply components in Chilean labour markets over three census periods (1982, 1992 and 2002). We aim to establish the way in which the changing distribution of economic activity in the Chilean landscape has shaped regional labour market outcomes in terms of employment, unemployment, labour force participation, migration, commuting and wages. Understanding of the interaction of these components is envisaged to provide the basis for future evidence-based, placed-oriented policy development.

NOTES

¹ We averaged fitness function values over the number of BTUs, U , to provide values between 0 and 1. This does not alter the final LMA delimitation.

² In this and the following cases, 'select an LMA (or BTU) with high (or low) numerical characteristic' refers to a three-way tournament selection; that is, three LMAs (or BTUs) are selected at random and the one with a higher (or lower) characteristic is chosen.

³ The CHIM geography differs from the current geography of Chile, which comprises 346 communes, 54 provinces and 15 regions. The number of spatial units increased after the division of existing communes, provinces and regions (INE, 2010).

⁴ In 2012, Chile conducted its 2010 round-Population and Housing census. However, data were not released. They were questioned due to problems of imputation of unobserved housing units, with no identification of imputed versus collected census data (Bianchini, Feeney, & Singh, 2013).

⁵ Three island communes – Easter Island, Juan Fernández and Antártica – were excluded due to their distant location and weak commuting interaction with continental Chile. Moreover, errors of

data transcription were identified during data processing. The corresponding census records of the communes of San Pedro de la Paz, San Pedro, Florida and La Florida were incorrectly coded. Similarity in their names appears to have been the problem. The communes identified report unexpectedly larger commuting flows to communes at very distant locations than towards neighbouring communes. Identification of the mistranscribed records was not possible so that they were corrected by redistributing commuting flows. For instance, 90% of the commuting flows from communes around San Pedro de la Paz that have San Pedro as a destination in the census database were assumed to be misrecorded and were reallocated to San Pedro de la Paz. The same procedure was applied to records that had a commune around La Florida as the place of origin and Florida as a commune of destination.

⁶ Each fitness function stands out for its own cohesion indicator. This indicator is a reflection of the fitness function being maximized.

⁷ Results are available from the authors upon request.

⁸ The analysis of the final GEA output revealed that of 12 misallocated BTUs three showed that their local self-containment would have improved through reassignment without compromising global and local LMA statistics. These three BTUs were thus reassigned: (1) Caldera from Diego de Almagro to Copiapo; (2) Loncoche from Panguipulli to Villarrica; and (3) Curacaví from Melipilla to Santiago.

⁹ In LMA delimitation there is a preference for detail (Casado-Díaz & Coombes, 2011): if there are two sets of valid LMAs, the set with more LMAs is usually preferred.

¹⁰ For brevity, we summarize our analyses in the paper. Details of the exhaustive analysis on each of the six LMA identified in the text are available from the authors upon request.

ACKNOWLEDGEMENTS

The authors acknowledge the comments of Professor Paul Elhorst, Dr Madeleine Hatfield, Professor Bernard Fingleton, four anonymous referees and Dr Thomas Sigler that helped to enhance the paper.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This paper was supported by the Spanish Ministry of Economy and Competitiveness [National Plan of R&D&I, grant number CSO2011-29943-C03-02] and the Chilean National Commission for Scientific and Technological Research through its human capital development programme.

SUPPLEMENTAL DATA

Companion software code developed by the authors for the improvements described in the paper can be accessed at <http://dx.doi.org/10.1080/17421772.2017.1273541>

ORCID

Francisco Rowe  <http://orcid.org/0000-0003-4137-0246>

REFERENCES

- Abalos, M., & Paredes, D. (2012). Una metodología para delimitar regiones urbanas funcionales (RUF) usando la conmutación a larga distancia: Evidencia empírica para Chile. *Documentos de Trabajo en Economía. Universidad Católica del Norte, Chile*. WP2012-10:1–26.
- Alba, E. (2002). Parallel evolutionary algorithms can achieve super-linear performance. *Information Processing Letters*, 82, 7–13.
- Alonso, M. P., Beamonte, A., Gargallo, P., & Salvador, M. (2015). Local labour markets delineation: An approach based on evolutionary algorithms and classification methods. *Journal of Applied Statistics*, 42(5), 1043–1063.
- Berdegú, J., Jara, B., Fuentealba, R., Tohá, J., Modrego, F., Schejtman, A., & Bro, N. (2011). Territorios funcionales en Chile. *Documento de Trabajo N102. Programa Dinámicas Territoriales Rurales. Rimisp Centro Latinoamericano para el Desarrollo Rural. Santiago, Chile*.
- Bianchini, R., Feeney, G., & Singh, R. (2013). Report of the International Commission on the 2012 Population and Housing Census of Chile, Instituto Nacional de Estadísticas.
- Blondel, V., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, 1–12.
- Brown, L., & Holmes, J. (1971). The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. *Journal of Regional Science*, 11, 57–72.
- Casado-Díaz, J. M., & Coombes, M. (2011). The delineation of 21st century local labour market areas: A critical review and a research agenda. *Boletín de la Asociación de Geógrafos Españoles*, 57, 7–32.
- Casado-Díaz, J. M., Martínez-Bernabeu, L., & Flórez-Revuelta, F. (2016). Automatic parameter tuning for functional regionalisation methods. *Papers in Regional Science*. doi:10.1111/pirs.12199
- Chakraborty, A., Beamonte, M. A., Gelfand, A. E., Alonso, M. P., Gargallo, P., & Salvador, M. (2013). Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets. *Computational Statistics & Data Analysis*, 58, 292–307.
- Coombes, M., & Bond, S. (2008). *Travel-to-work areas: The 2007 review*. London: Office for National Statistics.
- Coombes, M., Green, A., & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: The case of the 1984 travel-to-work areas revision in Britain. *Journal of the Operational Research Society*, 37, 943–953.
- Drobne, S., Konjar, M., & Lisec, A. (2010). Delimitation of functional regions of Slovenia based on labour market analysis. *Geodetski vestnik*, 54(3), 481–500.
- EUROSTAT, & Coombes, M. (1992). *Etude sur les zones d'emploi*. Luxembourg: Office for Official Publications of the European Communities.
- Farmer, C., & Fotheringham, A. (2011). Network-based functional regions. *Environment and Planning A*, 43, 2723–2741.
- Feria-Toribio, J. M., Casado-Díaz, J. M., & Martínez-Bernabeu, L. (2015). Inside the metropolis: The articulation of Spanish metropolitan areas into local labor markets. *Urban Geography*, 36(7), 1018–1041.
- Fjallstrom, P. (1998). Algorithms for graph partitioning: A survey. *Linköping Electronic Articles in Computer and Information Science*, 3(10), 1–34.
- Flórez-Revuelta, F., Casado-Díaz, J. M., & Martínez-Bernabeu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing*, 5, 10–21.
- Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104, 36–41.
- Fusco, G., & Cagliioni, M. (2011). Hierarchical clustering through spatial interaction data. The case of commuting flows in South-Eastern France. In *Computational science and its applications – ICCSA 2011* (pp. 135–151). London: Springer.
- Goodman, J. (1970). The definition and analysis of local labour markets: Some empirical problems. *British Journal of Industrial Relations*, 8, 179–196.

- Horst, B. (2005). Creación de nuevas regiones y comunas. Herramientas para su análisis. *Serie Informe Económico N162. Libertad y Desarrollo*: 1–35.
- Instituto Nacional de Estadísticas (INE). (2010). Código Único Territorial (CUT) 2010, ed. INE. Santiago, Chile: Santiago, Chile.
- Istituto Nazionale di Statistica (ISTAT). (2005). *I Sistemi Locali Del Lavoro 2001*. Roma: ISTAT.
- Istituto Nazionale di Statistica (ISTAT). (2015). *La nuova geografia del sistemi locali*. Roma: ISTAT.
- Klapka, P., Halás, M., Erlebach, M., Tonev, P., & Bednář, M. (2014). A multistage agglomerative approach for defining functional regions of the Czech Republic: The use of 2001 commuting data/ Vícestupňový aglomerační přístup k vymezení funkčních regionů České republiky: využití údajů o dojíždě z roku 2001. *Moravian Geographical Reports*, 22(4), 2–13.
- Kropp, P., & Schwengler, B. 2014. Three-step method for delineating functional labour market regions. *Regional Studies*, 50(3), 429–445.
- Lancichinetti, A., & Fortunato, S. (2011). Limits of modularity maximization in community detection. *Physical Review E*, 84, 1–9, article no. 066122.
- Landré, M. (2012). Geoprocessing journey-to-work data: Delineating commuting regions in Dalarna, Sweden. *ISPRS International Journal of Geo-Information*, 1, 294–314.
- Martínez-Bernabeu, L., & Casado-Díaz, J. M. (2016). Delineating zones to increase geographical detail in individual response data files: An application to the Spanish 2011 Census of population. *Moravian Geographical Reports*, 24(2), 26–36.
- Martínez-Bernabeu, L., Flórez-Revuelta, F., & Casado-Díaz, J. M. (2012). Grouping genetic operators for the delineation of functional areas based on spatial interaction. *Expert Systems with Applications*, 39, 6754–6766.
- Masser, I., & Brown, P. J. B. (1975). Hierarchical aggregation procedures for interaction data. *Environment and Planning A*, 7, 509–523.
- Newell, J. (2001). Scoping regional migration and its interaction with labour markets in New Zealand. *Occasional Paper 2001/2. New Zealand Department of Labour. Occasional Paper Series*: 1–52.
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69, 1–15, article no. 026113.
- Novak, J., Ahas, R., Aasa, A., & Silm, S. (2013). Application of mobile phone location data in mapping of commuting patterns and functional regionalization: A pilot study of Estonia. *Journal of Maps*, 9, 10–15.
- Organisation for Economic Co-operation and Development (OECD). (2002). *Redefining territories: The functional regions*. Paris: OECD.
- Organisation for Economic Co-operation and Development (OECD). (2009). *Territorial reviews: Chile 2009*. Paris: OECD.
- Office for National Statistics (ONS). (2015). *Methodology note on 2011 travel to work areas*. London: ONS.
- Prodromidis, P. (2008). Deriving labor market areas in Greece from commuting flows. *Athens, Greece. Centre for Planning and Economic Research (KEPE) No. 99, Athens Greece*: 5–28.
- Rowe, F. 2013a. *Spatial labour mobility in a transition economy: Migration and commuting in Chile* (Ph.D. thesis). School of Geography, Planning and Environmental Management, Faculty of Science, The University of Queensland, Brisbane, Australia.
- Rowe, F. 2014. *Micro and macro drivers of long-distance commuting in Chile: The role of spatial distribution of economic activities and population*. 53rd Western Regional Science Association. San Diego, United States, pp. 1–35.
- Rowe, F., & Bell, M. (2013). *Creating an integrated database for the analysis of spatial mobility in Chile* (Working Paper 02/2013). Queensland Centre for Population Research, School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, Australia.
- Rowe, F., & Bell, M. (Forthcoming 2017). The drivers of long-distance commuting in Chile: The role of spatial distribution of economic activities and population. In J. Poot & M. Roskrug (Eds.), *Regional science perspectives on population change and impacts in Asia and the Pacific*. Springer.
- Smart, M. W. (1974). Labour market areas: Uses and definition. *Progress in Planning*, 69, 238–353.
- Tolbert, C. M., & Sizer, M. K. (1987). *Labor market areas for the United States*. Washington, DC: Economic Research Service, U.S. Dept. Agr.

- Tomassini, M. (2005). *Spatially structured evolutionary algorithms*. New York: Springer.
- Watts, M. (2013). Assessing different spatial grouping algorithms: An application to the design of Australia's new statistical geography. *Spatial Economic Analysis*, 8, 92–112.
- Zegras, C., & Gakenheimer, R. (2000). Urban growth management for mobility: The case of the Santiago. Chile Metropolitan Region: Report prepared for the Lincoln Institute of Land Policy and the MIT Cooperative Mobility Program.